Brain and Cognition 124 (2018) 29–36

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Selective impairments in components of affective prosody in neurologically impaired individuals

Amy Wright[a], Sadhvi Saxena[a], Shannon M. Sheppard[a], Argye E. Hillis[a,b,c,*]

[a] Johns Hopkins University School of Medicine, Department of Neurology, USA
[b] Johns Hopkins University School of Medicine, Department of Physical and Medicine & Rehabilitation, USA
[c] Johns Hopkins University, Department of Cognitive Science, USA

ARTICLE INFO

Keywords:
Prosody
Emotions
Social communication
Stroke

ABSTRACT

The intent and feelings of the speaker are often conveyed less by what they say than by how they say it, in terms of the affective prosody – modulations in pitch, loudness, rate, and rhythm of the speech to convey emotion. Here we propose a cognitive architecture of the perceptual, cognitive, and motor processes underlying recognition and generation of affective prosody. We developed the architecture on the basis of the computational demands of the task, and obtained evidence for various components by identifying neurologically impaired patients with relatively specific deficits in one component. We report analysis of performance across tasks of recognizing and producing affective prosody by four patients (three with right hemisphere stroke and one with frontotemporal dementia). Their distinct patterns of performance across tasks and quality of their abnormal performance provides preliminary evidence that some of the components of the proposed architecture can be selectively impaired by focal brain damage.

1. Introduction

Here we propose a preliminary architecture of the perceptual, cognitive, and motor processes underlying recognition and production of affective prosody (variations in pitch, volume, rate, and rhythm of speech to convey emotion in verbal language) on the basis of the computational demands of the task. We obtained support for this architecture through analysis of performance across affective prosody tasks by neurologically impaired participants with distinct impairments in the proposed cognitive architecture. The patterns of performance of four individual cases with relatively selective deficits in affective prosody provide preliminary support for the hypothesis that at least some of the perceptual and cognitive processes underlying affective prosody recognition and production are dissociable.

Affective prosody refers to variations in pitch, loudness, rate, and rhythm (pauses, stress, duration of components) of speech to convey emotions. Many studies have shown that recognition and/or production of affective prosody can be impaired by focal (or diffuse) brain damage (Pell, 2002, 2006), most notably due to right hemisphere stroke (Ross & Mesulam, 1979; Ross & Monnot, 2008; Tippett & Ross, 2015), frontotemporal dementia (Dara et al., 2013; Rankin et al., 2009; Rankin, Kramer, & Miller, 2005), Parkinson's disease (Péron et al., 2015), schizophrenia (Dondaine et al., 2014), and other neurological diseases (Bais, Hoekert, Links, Knegtering, & Aleman, 2010; Kipps, Duggins, McCusker, & Calder, 2007). Several studies have demonstrated dissociations between recognition and production of affective prosody impairment, and some have shown that repetition or mimicking of affective prosody can be disproportionately impaired or spared (Ross & Monnot, 2008). However, few studies have attempted to identify distinct perceptual, cognitive, and motor mechanisms that underlie affective prosody recognition and production [but see (Bowers, Bauer, & Heilman, 1993)].

A number of investigators have proposed neural systems underlying either recognition or production of affective prosody. For example, a neurobiological model of emotional information processing has recently been proposed (Brück, Kreifelts, & Wildgruber, 2011), on the basis of a review of fMRI studies and lesion studies of prosody. In this model, affective prosody is initially processed in the thalamus, which connects to both an explicit pathway and an implicit pathway. The explicit pathway includes extraction of acoustic cues in the middle part of superior temporal gyrus (mSTG), integration of the acoustic cues into a single percept in posterior superior temporal gyrus (pSTG), and evaluation of the percept in dorsolateral prefrontal cortex and orbitofrontal cortex. The implicit pathway projects directly to the amygdala, insula, nucleus accumbens, and medial frontal cortex to induce an emotional response. The authors add that there is evidence, primarily from

* Corresponding author at: Department of Neurology, Phipps 446, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA.
E-mail address: argye@jhmi.edu (A.E. Hillis).

https://doi.org/10.1016/j.bandc.2018.04.001
Received 10 March 2018; Received in revised form 13 April 2018; Accepted 15 April 2018

Parkinson's disease and fMRI that the basal ganglia also have a role in recognition of affective prosody, but the role may be a more general one of processing timing of acoustic information or working memory. A neurobiological model that shares some features is a three stage model proposed by Schirmer and Kotz (2006). This model includes: (1) acoustic analysis in bilateral auditory processing areas; (2) integration of emotionally significant acoustic information into an emotional 'gestalt' along the ventral stream from STG to anterior superior temporal sulcus (STS), which seems to be lateralized to the right hemisphere; and (3) higher-order cognitive processes that yield explicit evaluative judgments of the derived emotional gestalt mediated by the right inferior gyrus and orbitofrontal cortex. Reviews of fMRI studies of affective prosody comprehension have found support for a right-lateralized network including the STG and STS, as well as inferior frontal gyrus (Kotz, Meyer, & Paulmann, 2006; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Although the right hemisphere may have a dominant role in affective prosody, there is clearly a role of the left hemisphere in processing prosody as well, as shown by fMRI studies (Kotz et al., 2003; Wildgruber et al., 2006) and lesion studies in which left hemisphere strokes result in impaired production and/or recognition of prosody in speech (Baum & Pell, 1997, 1999; Cancelliere & Kertesz, 1990; Schlanger, Schlanger & Gerstman, 1976). Furthermore, integration of information from the two hemispheres, via the corpus callosum, is essential for accurate processing of affective prosody (Paul, Van Lancker-Sidtis, Schieffer, Dietrich & Brown, 2003). Previous investigations have also provided evidence that elements of prosody perception are differently lateralized, with pitch processed mostly in the right hemisphere, while duration and intensity are processed mainly in the left hemisphere (Van Lancker & Sidtis, 1992; Zatorre, Belin, & Penhune, 2002). In all of the studies that support these neurobiological models, there has been no clear evidence that emotional valence or type of emotion significantly influences the areas that are engaged.

Neurobiological models of *generation* of affective prosody during speech production have implicated a right-lateralized network (Riecker, Wildgruber, Dogil, Grodd, & Ackermann, 2002), bilateral perisylvian network (Aziz-Zadeh, Sheng, & Gheytanchi, 2010), or a critical role of the basal ganglia (Cancelliere & Kertesz, 1990; Van Lancker & Sidtis, 1992). A recent study found support for all three of these networks/regions in different aspects of prosody generation (Pichon & Kell, 2013). In this fMRI study, during the preparatory phases of prosody generation, there was increased ipsilateral connectivity between right ventral and dorsal striatum and between the striatum and the anterior STG, temporal pole, and right anterior insula. Additionally, there was increased connectivity between right dorsal striatum and right inferior frontal gyrus (Brodmann area 44), and left orbitofrontal cortex and inferior temporal gyrus. During the execution phase, connectivity increased between the right ventral striatum and dorsal striatum, and between dorsal striatum and bilateral inferior frontal gyrus and right STG, but also between right dorsal striatum and right anterior hippocampus and bilateral amygdala.

While these neurobiological models are useful for understanding areas of the brain engaged in affective prosody recognition and/or production, they are underspecified in terms of the cognitive processes that underlie the ubiquitous and socially imperative tasks of recognizing and conveying emotion through prosody. How is it that everyone in the room is able to recognize when a speaker is angry, sorrowful, afraid, or delighted, not by what they say, but how they say it? Likewise, how is it that we all convey our own emotions not so much by what we say, but how we say it, in such a way that the emotion is universally understood (even if the listener does not share the language)?

To develop a cognitive architecture of the representations and processes underlying recognition and expression of affective prosody, we can start by considering the computational demands of recognizing and expressing emotions through changes in the acoustic features of speech. To recognize emotion in another's speech, it is first essential to parse and analyze the paralinguistic features of the utterance. This acoustic analysis requires recognition of differences in pitch (frequency), loudness (intensity), rate, and rhythm (stress, pauses, and duration of various segments). Then, it is necessary to match a set of acoustic features with an emotion. To do this, we need access to an abstract representation of what "angry", "sad", "happy" and so on sound like. These abstract representations of acoustic characteristics that convey emotion (ARACCE) are shared by speakers of a language or culture. The ARACCE (whether stored or computed on-line) would specify the acoustic features (e.g., low frequency, high intensity, rapid rate and their interactions) of each emotion (e.g., anger), and allow access to the semantic representation, or meaning, of that emotion. Thus, the ARACCE is comparable to the lexical orthographic representation for reading and spelling, in that it mediates between the semantic representation and input or output processes. The semantic representation would specify an autonomic response (i.e., the so-called "fight or flight" or sympathetic response or parasympathetic response) as well as the meaning and valence, and would overlap with semantic representations of other emotions. For example, disappointed and heartbroken share aspects of meaning, but vary in valence. To express an emotion (e.g. anger) through prosody, one would need to access the abstract representation of the acoustic features of anger (the ARACCE) from the semantic representation of anger. Then, one would need to convert the ARACCE to motor programs for producing these acoustic features (e.g. changes in length and tension of the vocal folds by constriction or relaxation of cricothyroid and other laryngeal and respiratory muscles, changes in rate of movement of the lips, tongue, jaw, palate, and so on). Finally, one would need to *implement* these motor programs during the complex act of speaking. These proposed mechanisms are schematically represented in Fig. 1. Although we have discussed these mechanisms as though they are activated serially, it is likely that some operate in parallel during listening or production of speech. Also, there may be feedforward or feedback interactions between levels that are not depicted here.

In this study, we provide empirical evidence for some of the proposed levels of processing, by showing that they can be relatively selectively disrupted by brain damage. Participants with acute right hemisphere stroke or frontotemporal dementia (FTD) were tested on a battery of tests designed to assess processing at each proposed level. We selected these two populations because impaired prosody is common after both right hemisphere stroke (Dara, Bang, Gottesman, & Hillis, 2014; Ross & Mesulam, 1979) and FTD (Dara et al., 2013; Phillips, Sunderland-Foggio, Wright, & Hillis, 2017; Rankin et al., 2009). Unlike individuals with autism or schizophrenia, it is possible to identify the region of infarct or focal atrophy associated with the impaired level of processing.

In this ongoing study, many participants showed no impairment at any level and others had deficits at several levels of processing or insufficient testing to determine the level(s) of processing that were impaired. Nevertheless, we identified four participants whose performance across tasks and acoustic characteristics of speech can be accounted for by assuming selective disruption of a single level.

2. Methods

2.1. Participants

A total of four neurologically impaired individuals were included in this case series, including one with FTD and three participants with acute, right hemisphere stroke. All participants were recruited from the Johns Hopkins Hospital stroke service or the cognitive disorders clinic; all were right-handed. They were selected from a total of nine patients with FTD and 22 patients with acute ischemic right hemisphere stroke who completed the Affective Prosody Battery. Additionally, 60 healthy, age- and education-matched neurotypical controls were tested on one or more of the tests (see numbers for each test in Table 1). Exclusion